

NCI Cancer Genomics Cloud Pilots

Bringing data and computation together to create knowledge that accelerates cancer research and enables precision medicine.

The advent of large-scale scientific programs that use next-generation sequencing technology has resulted in an enormous growth of biomedical data. The volume of data makes access and analysis difficult in terms of both storage and computing capability.

The Cancer Genomics Cloud Pilots are designed to explore innovative methods for accessing and computing on large scale cancer genomic data. They bring analytics and data from The Cancer Genome Atlas (TCGA) together on a single platform through a set of data repositories with co-located computational capacity and an Application Programming Interface (API) that provides secure data access. In this model, applications are brought to the data, rather than bringing the data to the applications.

- Access and analyze 11,000 TCGA samples without having to download data
- Upload your own data for analysis

Data



- Perform large scale analysis using the elastic compute power of commercial cloud platforms

Compute



- dbGaP-authorized users can access controlled TCGA data
- Systems meet strict Federal security guidelines

Security



The Cancer Genomics Cloud Pilots are now available!

We invite you to try out the Cloud Pilots. Cloud credits to support your evaluation activities may be available. Please contact the individual Cloud Pilots for details.

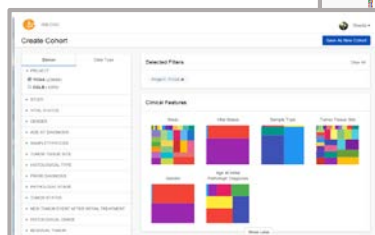
Your feedback is important to NCI's future planning for cloud computing. Please complete our survey once you have gained experience with one or more of the systems. www.surveymonkey.com/r/NCIGenomicsCloudPilots



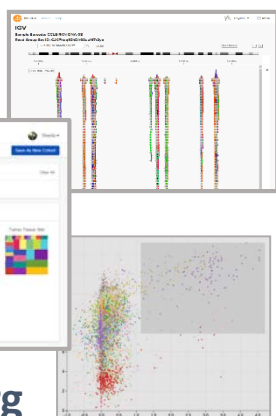
For more information about the Cancer Genomics Cloud Pilot Program, please visit: cbiit.nci.nih.gov/cloudpilots or email ncicloud@nih.gov with your questions, ideas, and suggestions.

The goals of the NCI Cloud Pilots are to democratize access to NCI-generated genomic and related data, and to create a cost-effective way to provide scalable computational capacity to the cancer research community.

Institute for Systems Biology

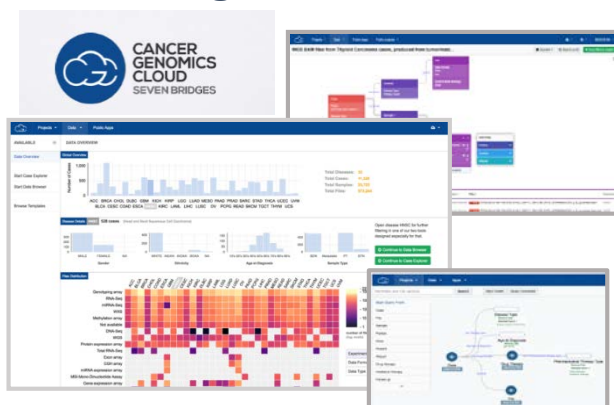


www.isb-cgc.org



The **Institute for Systems Biology (ISB) Cloud** provides interactive and programmatic access to data, leveraging many aspects of the Google Cloud Platform. The **interactive ISB-CGC web-app** allows scientists to interactively define and compare cohorts, examine underlying molecular data for specific genes or pathways of interest, and share insights with collaborators. For computational users, programmatic interfaces and GCP tools such as **BigQuery**, **Genomics**, and **Compute Engine** allow users to perform complex queries from R or Python scripts, or run Dockerized workflows on sequence data available in cloud storage.

Seven Bridges Genomics



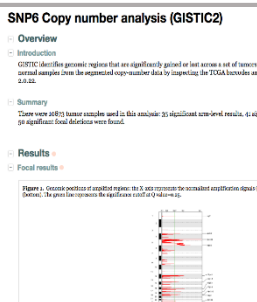
www.cancergenomicscloud.org

Seven Bridges Genomics Cancer Genomics Cloud enables researchers to collaborate on the analysis of large cancer genomics datasets in a secure, reproducible, and scalable manner. A **rich query system** allows researchers to find the exact data of interest and combine it with their own private data. Native implementation of the **Common Workflow Language specification** makes it easy for developers, analysts, and bench biologists to deploy, customize and run reproducible analysis methods to learn from genomics data faster.

Broad Institute



www.firecloud.org



Broad Institute FireCloud is modeled after their **Firehose analysis infrastructure** and facilitates collaboration and provides a robust, scalable platform accessible to the community at-large. Using the elastic compute capacity of Google Cloud, FireCloud empowers analysts, tool developers, and production managers to perform large-scale analysis, engage in data curation, and store or publish results. Users can upload their own analysis methods and data to workspaces or run the **Broad's best practice tools and pipelines** on pre-loaded data.